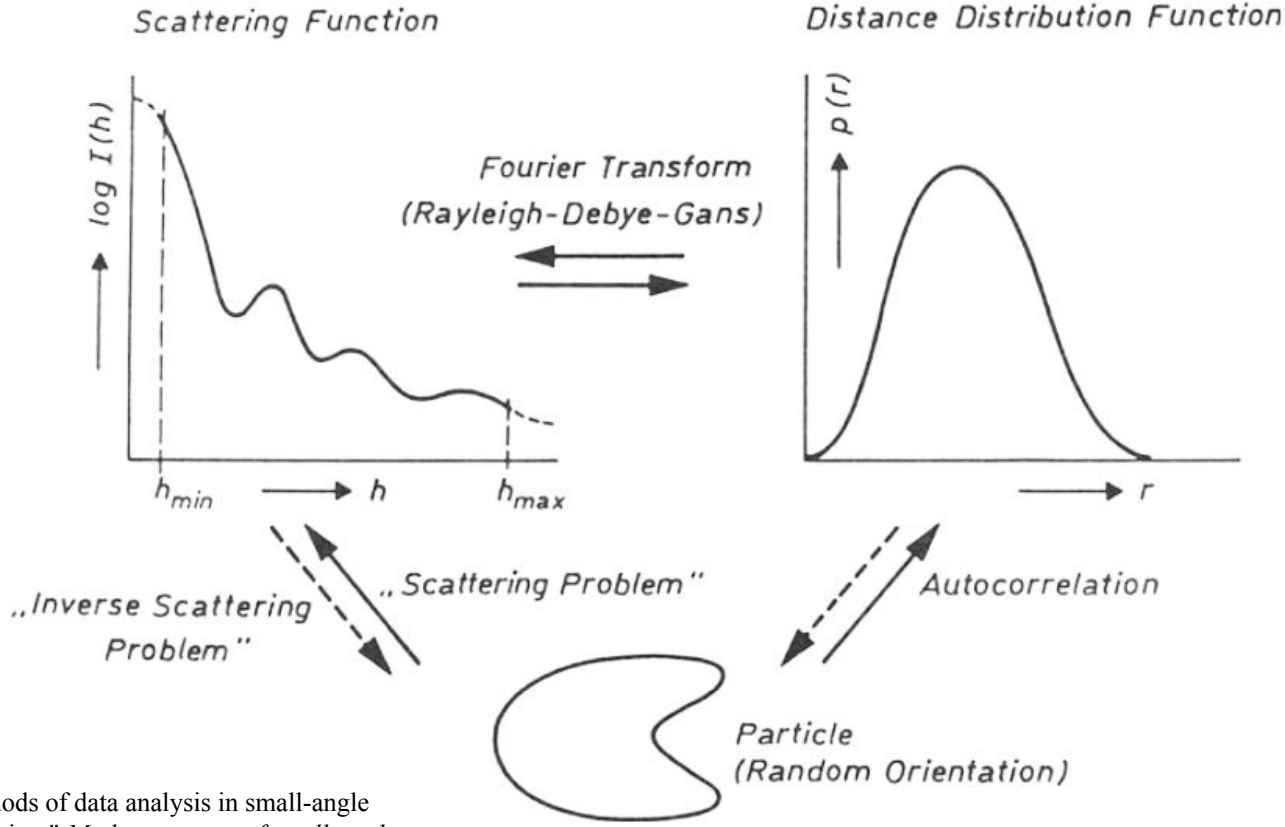
Molecular Dynamics Simulations, Small-Angle Scattering and The Inverse Scattering Problem

Marie Lycksell

NEUTRONS FOR SOCIETY

FASEM research school, Lund, 2024-03-12

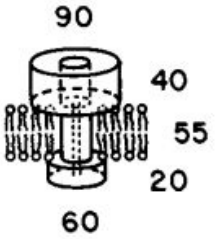
The inverse scattering problem



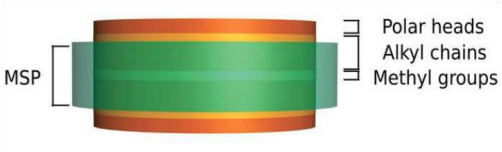
Glatter, O. "Modern methods of data analysis in small-angle scattering and light scattering." *Modern aspects of small-angle scattering*. Dordrecht: Springer Netherlands, 1995. 107-180.

Hypotheses for the particle shape

Geometry + Scattering length densities



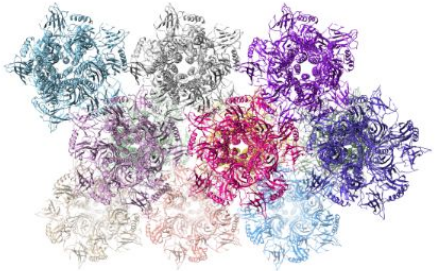
Wise, D. S., Karlin, A., and Schoenborn, B. P. *Biophysical journal* 28.3 (1979): 473-496.



Orioli, S., Henning Hansen, C. G., and Arleth, L. *Acta Crystallographica Section D: Structural Biology* 77.2 (2021): 176-193.

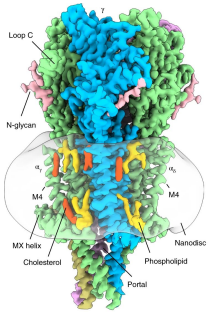
Models of atom locations

X-ray diffraction



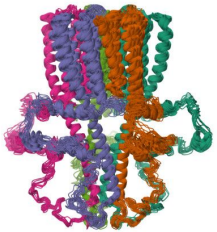
Hu, H., et al. *Proceedings of the National Academy of Sciences* 117.24 (2020): 13437-13446.

CryoEM



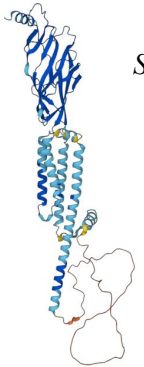
Rahman, M. M., et al. *Nature structural & molecular biology* 29.4 (2022): 386-394.

NMR



Bondarenko, V., et al. *Nature communications* 13.1 (2022): 793.

Structure prediction



https://alphafold.ebi.ac.uk/entry/P36544

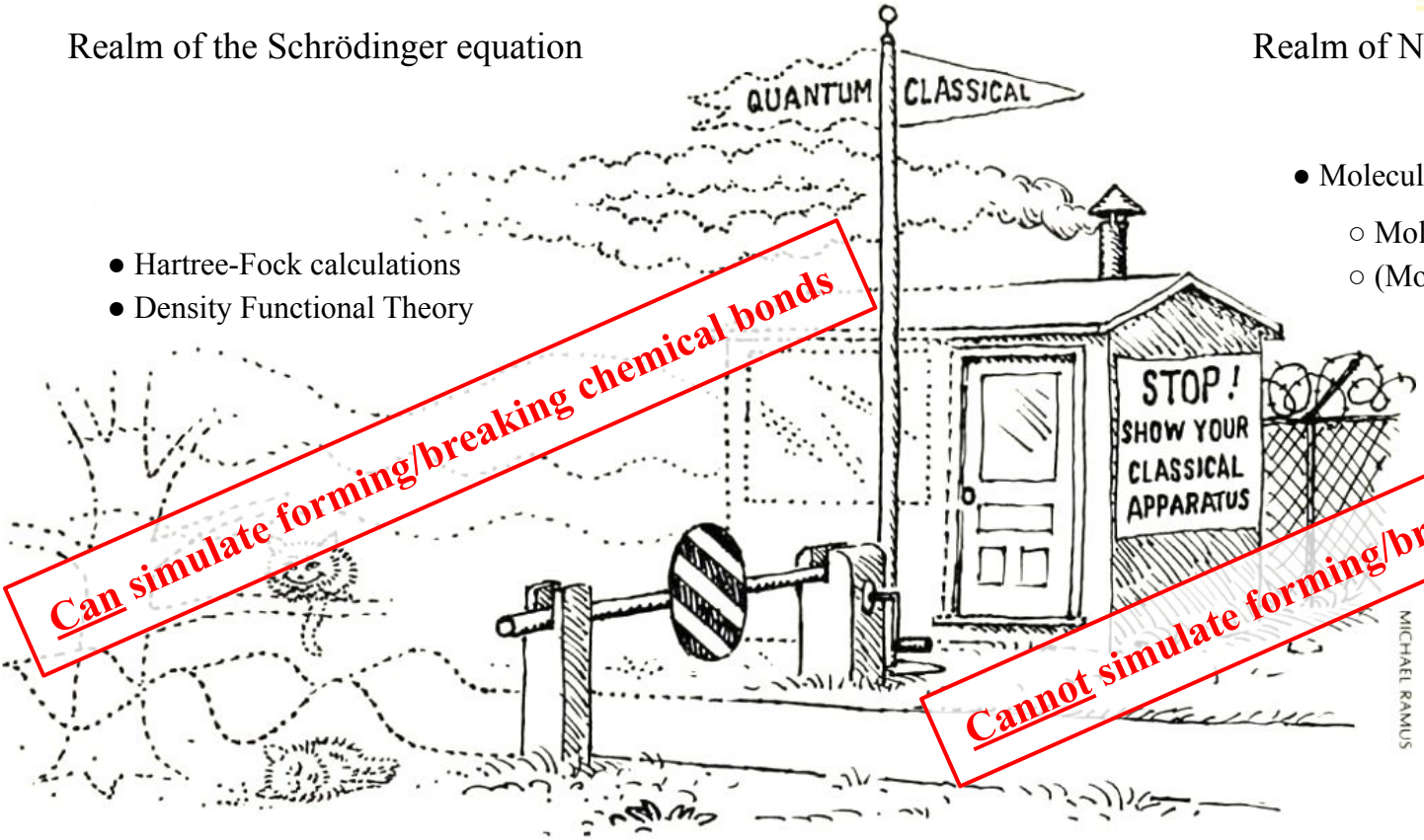
Molecular simulations

Realm of the Schrödinger equation

- Hartree-Fock calculations
- Density Functional Theory

Realm of Newton's equations

- Molecular Mechanics
  - Molecular Dynamics simulations
  - (Monte Carlo simulations)



Can simulate forming/breaking chemical bonds

Cannot simulate forming/breaking chemical bonds

Image from: Zurek, W. H. "Decoherence and the transition from quantum to classical." *Physics today* 44.10 (1991): 36-44.

Molecular Dynamics simulations vs. Monte Carlo simulations

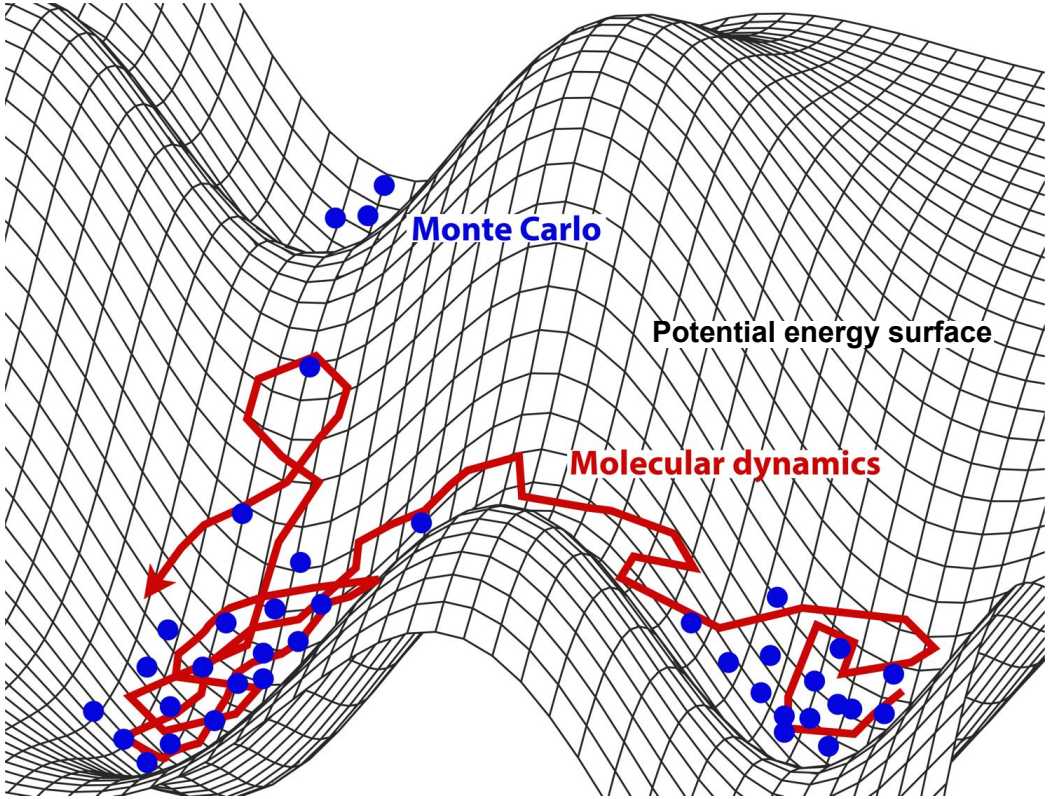
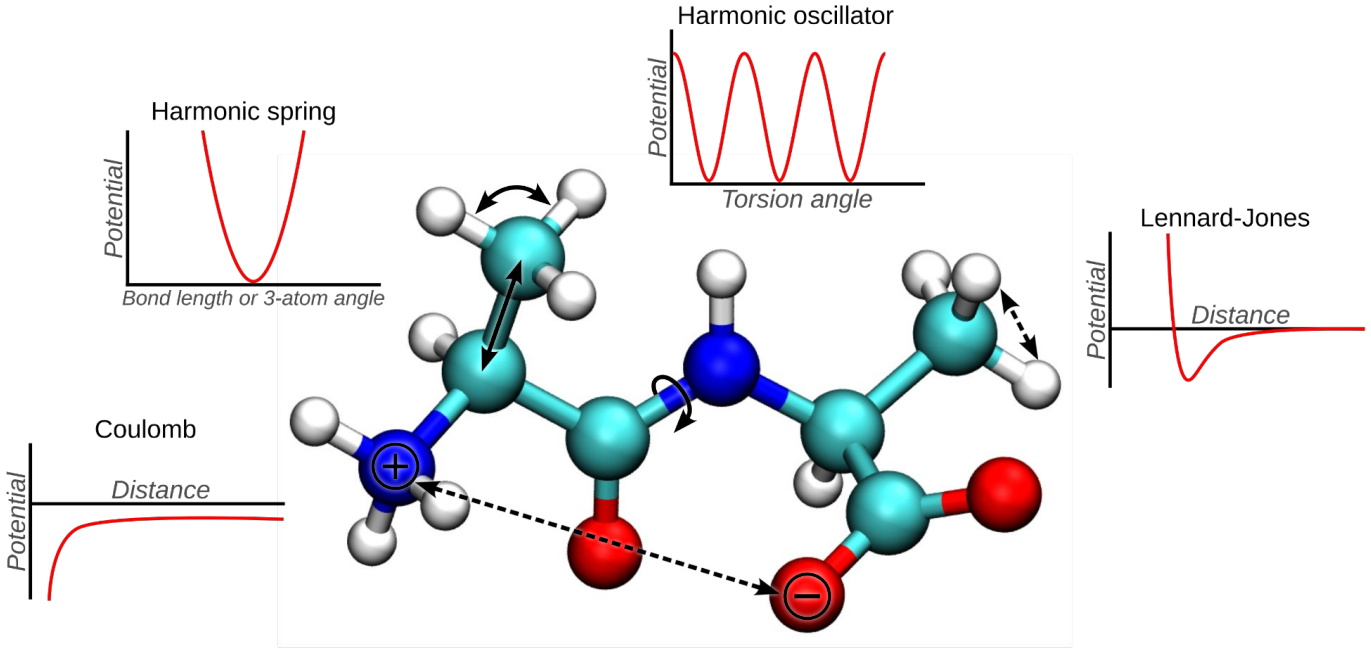


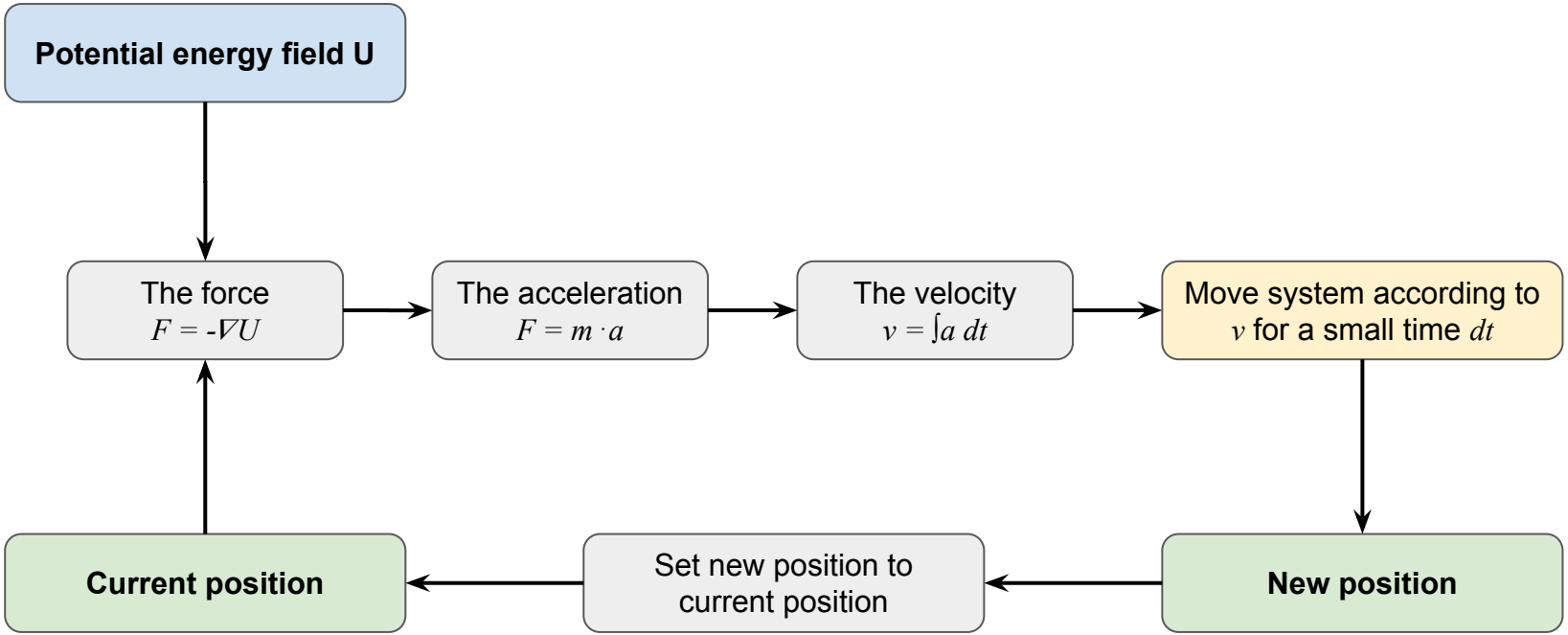
Image from:
https://commons.wikimedia.org/wiki/File:Sampling_in_Monte_Carlo_and_molecular_dynamics.png

The Force Field: Molecular Mechanics Potential Energy Surfaces

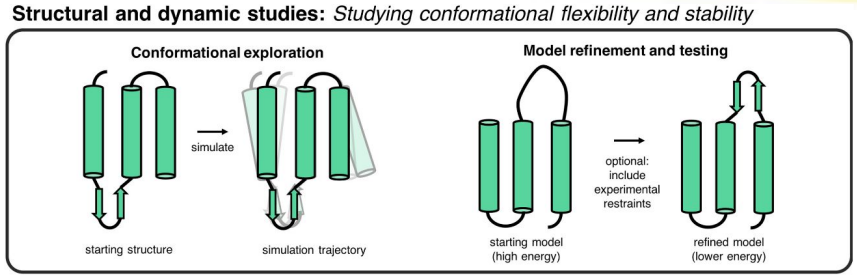
$$U = \underbrace{U_{bond} + U_{angle} + U_{torsion}}_{\textit{Bonded interactions}} + \underbrace{U_{vdw} + U_{electrostatic}}_{\textit{Non-bonded interactions}}$$

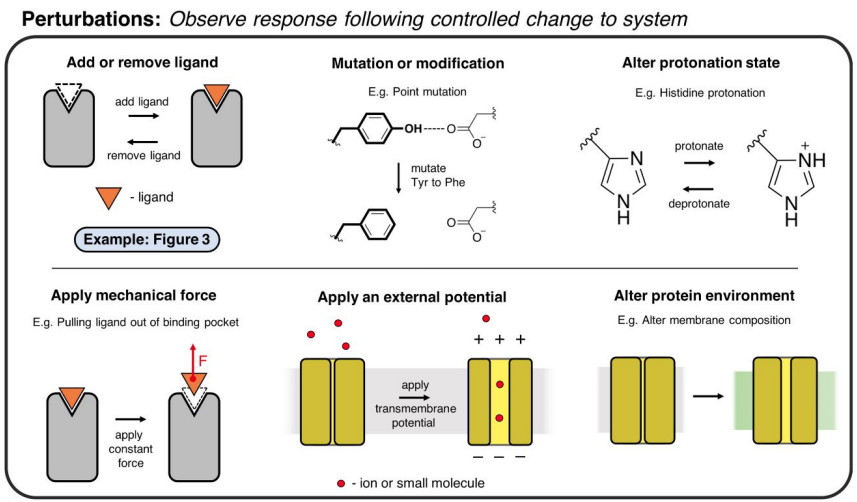


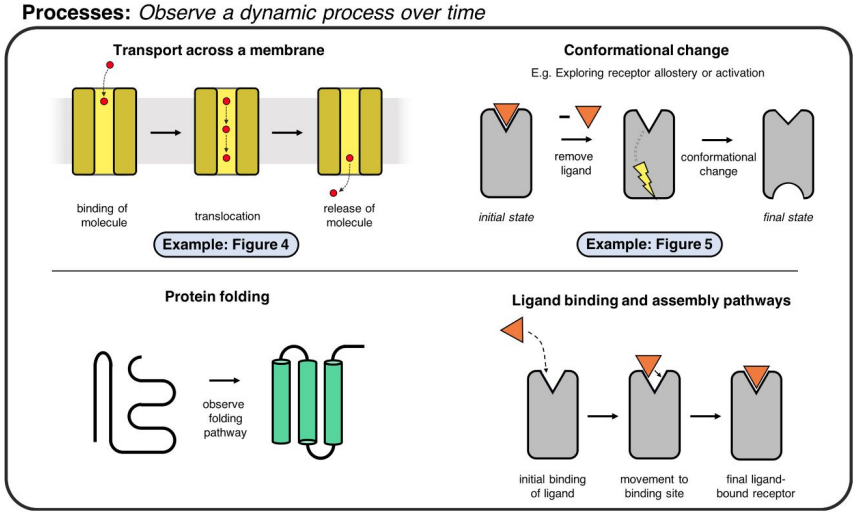
The Molecular Dynamics simulation loop



MD simulations can be used to study many types of questions







Hollingsworth, S. A., and Dror, R. O. . "Molecular dynamics simulation for all." *Neuron* 99.6 (2018): 1129-1143.

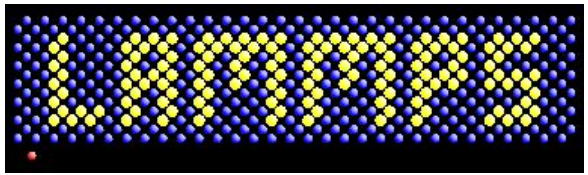
A non-exhaustive list of software for Molecular Dynamics simulations

FAST. FLEXIBLE. FREE.
GROMACS

NAMD
Scalable Molecular Dynamics
VMD
Visual Molecular Dynamics

CHARMM
Chemistry at HARvard Macromolecular Mechanics
CHARMM-GUI
Effective Simulation Input Generator and More

Amber



High performance computing

or

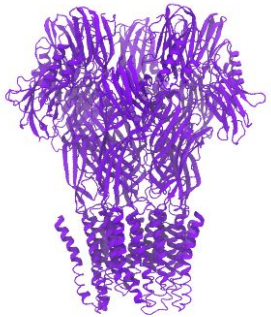
EuroHPC
Joint Undertaking

PARTNERSHIP FOR ADVANCED
COMPUTING IN EUROPE

NAISS

GENCI
Le calcul intensif au service de la connaissance

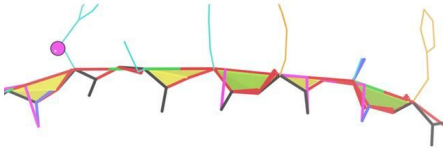
Setting up a system for MD simulations

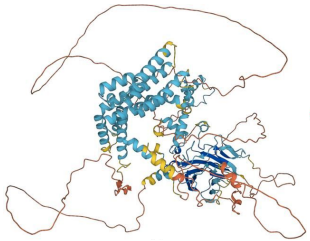


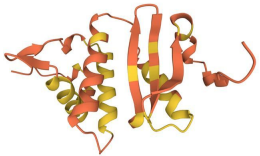
Atomistic model

- Experimental structure
  - Add missing residues

- Predicted structure
  - Remove or re-model “barbed wire” in AlphaFold models

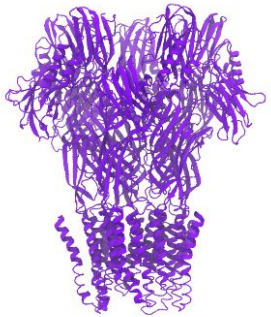






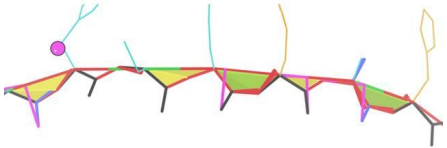
Richardson, J. S., et al. *Acta Crystallographica Section D: Structural Biology* 79.12 (2023).

Setting up a system for MD simulations

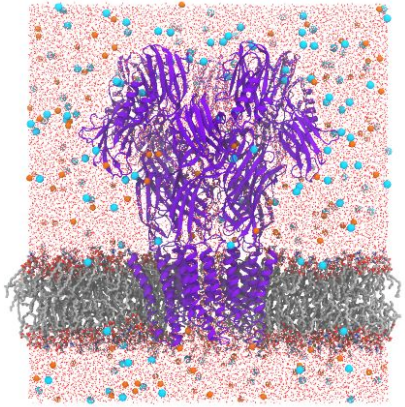


Atomistic model

- Experimental structure
  - Add missing residues
- Predicted structure
  - Remove or re-model “barbed wire” in AlphaFold models



Richardson, J. S., et al. *Acta Crystallographica Section D: Structural Biology* 79.12 (2023).



Add additional components

- Lipid bilayers
- Ligands
- Etc.

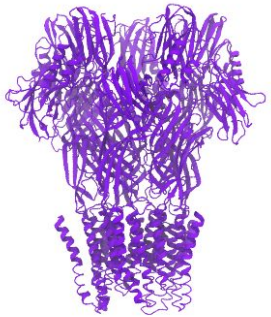
Solvate the system

- Add water
- Add ions
  - Neutralise system charge
  - Desired concentration

Topology file

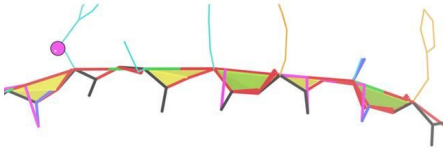
- Lists what is in the system
- Describes the connectivity of molecules in the system
  - PDB files lists atom coordinates, but not which atoms are covalently bound
  - Topology files contain constant attributes of atoms, not dynamic attributes like positions
- Lists which force field files to use

Setting up a system for MD simulations

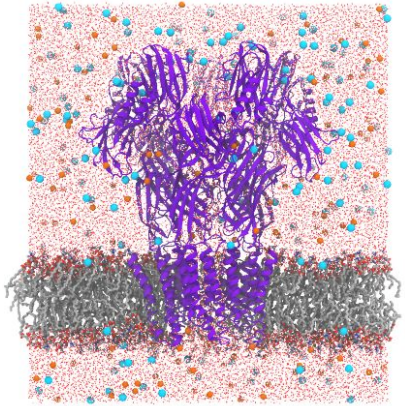


Atomistic model

- Experimental structure
  - Add missing residues
- Predicted structure
  - Remove or re-model “barbed wire” in AlphaFold models



Richardson, J. S., et al. *Acta Crystallographica Section D: Structural Biology* 79.12 (2023).

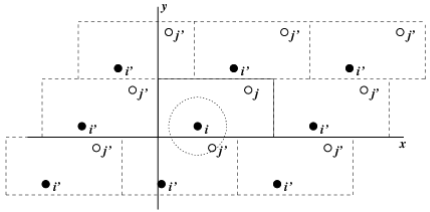


Add additional components

- Lipid bilayers
- Ligands
- Etc.

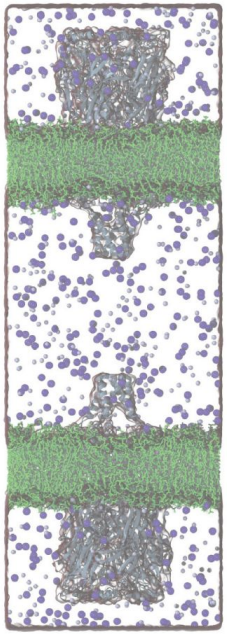
Solvate the system

- Add water
- Add ions
  - Neutralise system charge
  - Desired concentration



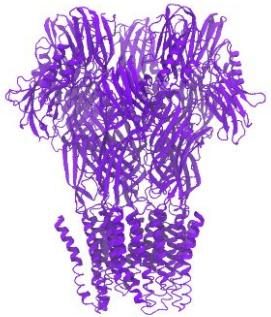
Periodic boundary conditions

- The simulation box is treated as a repeating unit
  - Neighbours itself to avoid edge effects
- Box must be large enough
  - The protein shouldn’t sense its periodic image (unless simulating a crystal)
- Take into account if regions with different concentrations are desired



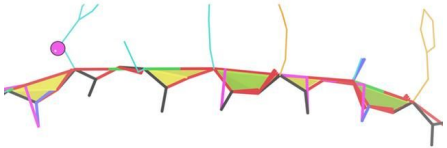
Zhuang, Y., et al. *PNAS* 119.43 (2022): e2208081119.

Setting up a system for MD simulations

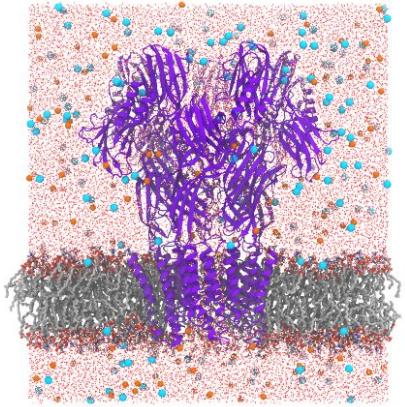


Atomistic model

- Experimental structure
  - Add missing residues
- Predicted structure
  - Remove or re-model “barbed wire” in AlphaFold models



Richardson, J. S., et al. *Acta Crystallographica Section D: Structural Biology* 79.12 (2023).



Add additional components

- Lipid bilayers
- Ligands
- Etc.

Solvate the system

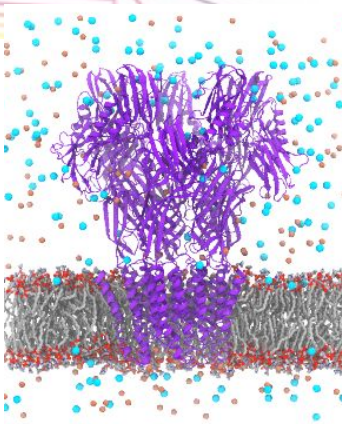
- Add water
- Add ions
  - Neutralise system charge
  - Desired concentration

Energy minimization

- Relax the system to a (local) minima
  - Removes steric clashes
  - Avoids inappropriate geometry

Equilibration

- Bring the system up to temperature and pressure
- Optimize solvent with respect to the solute
- NVT
  - Stabilize temperature
- NPT
  - Stabilize pressure
  - Gradual release of restraints
    - Heavy atoms
    - Backbone atoms
    - C-alpha atoms

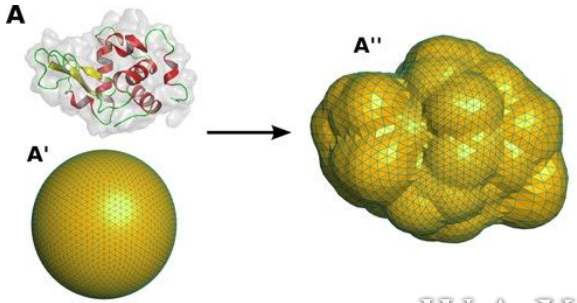


Production run

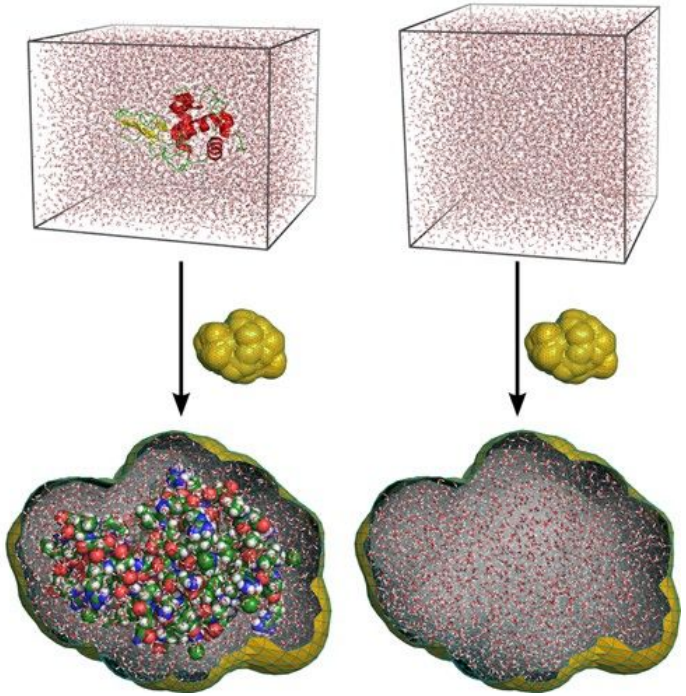
- The full length simulation
- No restraints (unless part of the intent is to have them)
- Launch replicas
  - Often better to have four trajectories of e.g. 500 ns than just one of 2 µs

MD simulations can be used to describe the hydration layer

- The hydration layer is typically denser than the bulk solvent, and has internal structure
  ⇒ Must be taken into account when fitting SAXS data
  - Density of the hydration layer is a common parameter in fitting software
  - Explicit solvent MD simulations can be used instead to capture the hydration layer
- The WAXSIS web server by Jochen S. Hub et al. facilitates using MD to fit SAXS data.

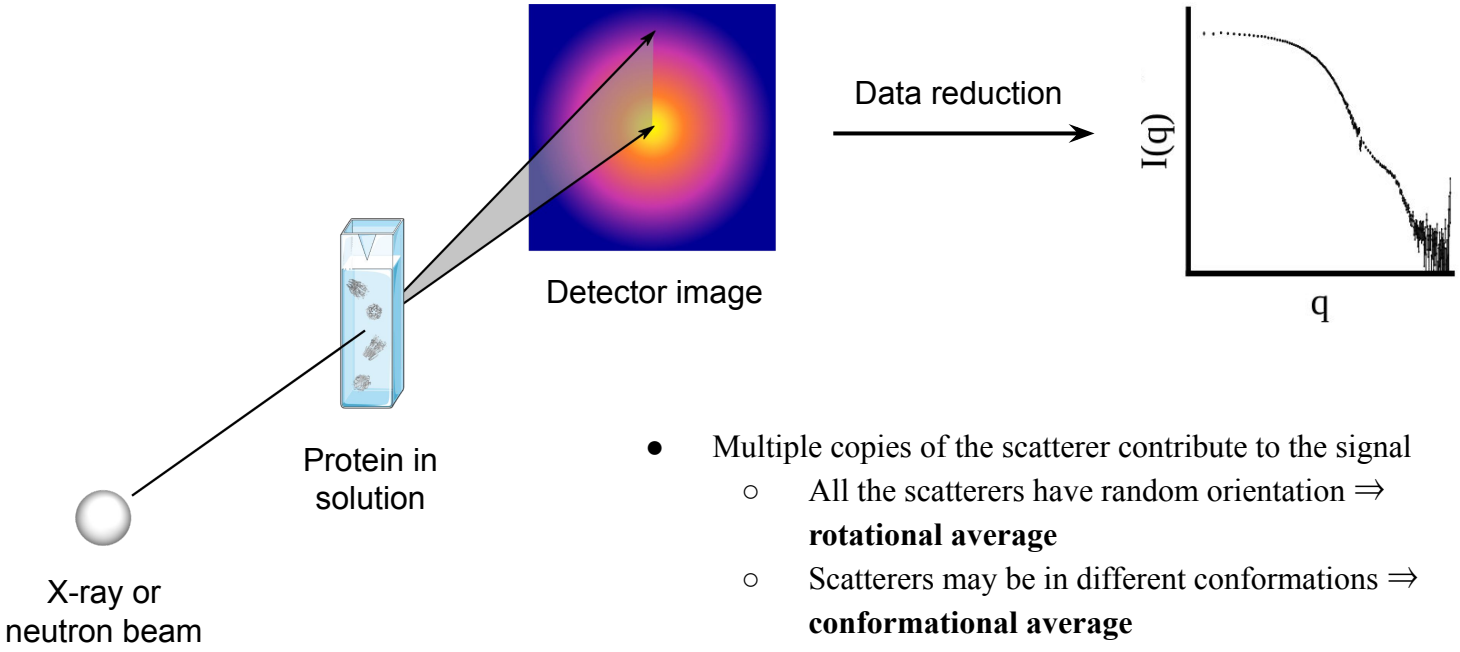


WAXSiS



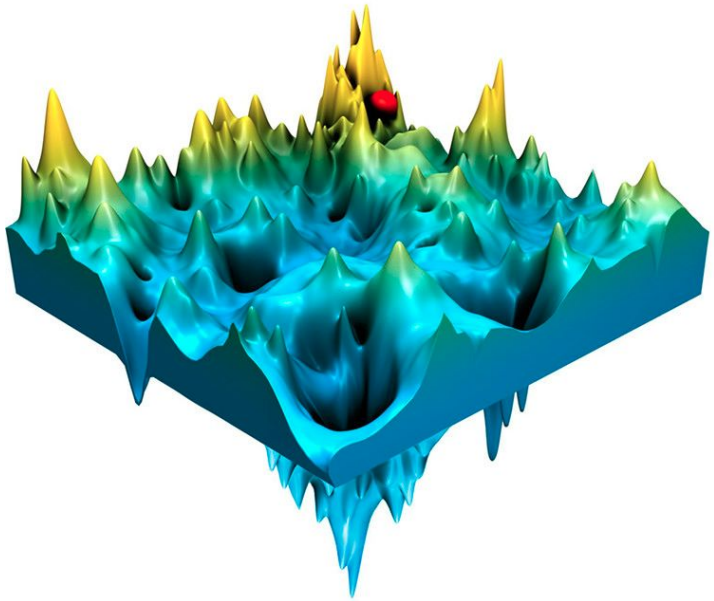
Knight, C. J., and Hub, J. S. "WAXSiS: a web server for the calculation of WAXS/WAXS curves based on explicit-solvent molecular dynamics." *Nucleic acids research* 43.W1 (2015): W225-W230.

SAS yields information about the population average



- Multiple copies of the scatterer contribute to the signal
  - All the scatterers have random orientation ⇒ **rotational average**
  - Scatterers may be in different conformations ⇒ **conformational average**
- Data collection takes time (especially with neutrons) and conformation may change ⇒ **time average**

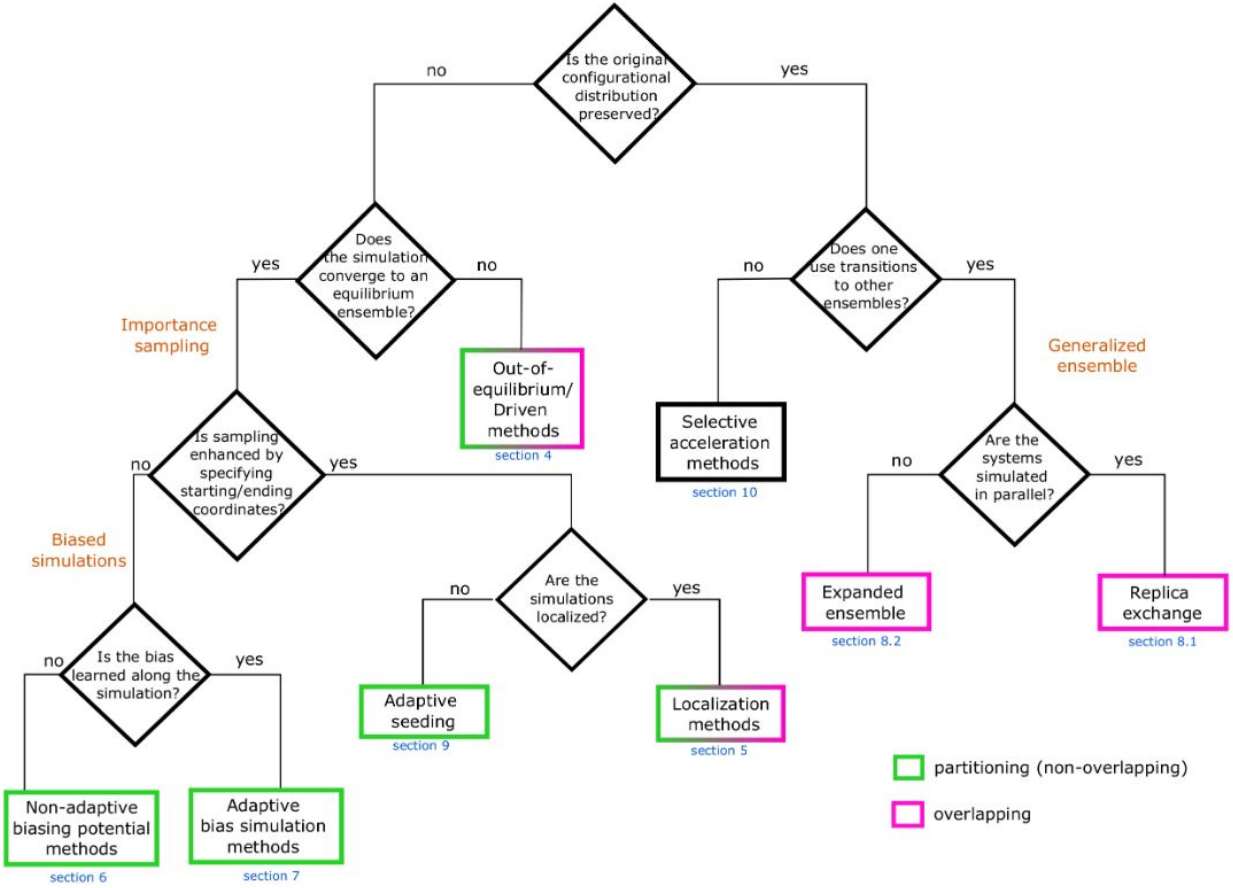
The sampling problem



- MD simulations are prone to sample local minima
- It is common that the conformational changes of interest don’t happen spontaneously during the timescales accessible to “vanilla” MD

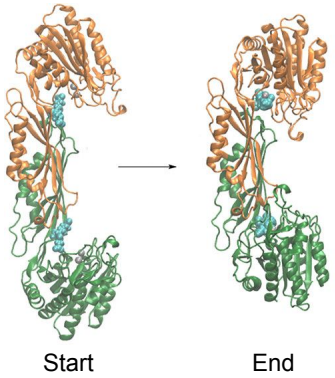
Image from: Kmiecik, S., et al. "Coarse-grained protein models and their applications." *Chemical reviews* 116.14 (2016): 7898-7936.

Many, many enhanced sampling schemes exist



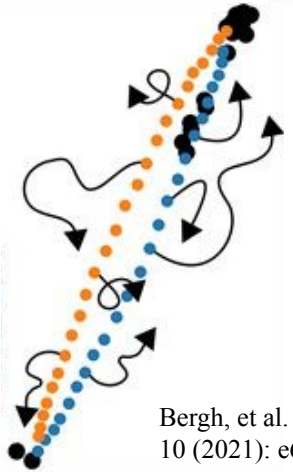
Hénin, J., et al. "Enhanced sampling methods for molecular dynamics simulations." *arXiv preprint arXiv:2202.04164* (2022).

A few ways to access more of conformational space



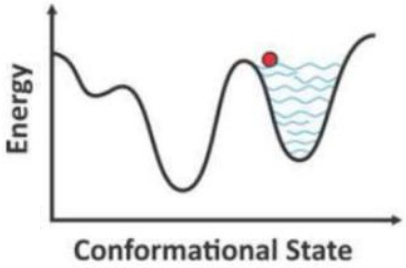
Pull the starting structure towards a target
- *Steered MD*

Kochert, et al. *Biochemistry* 60.12 (2021): 908-917.



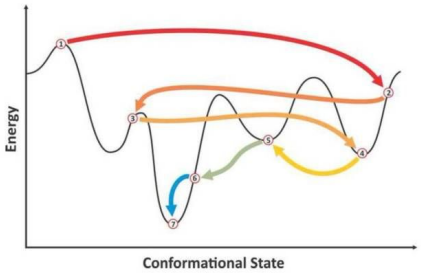
Extrapolate between known states, launch simulations from many intermediates
- *The string method with swarms of trajectories*

Bergh, et al. *Elife* 10 (2021): e68369.



Add energy penalties to visited conformations
- *Metadynamics*
- *AWH*

Bernardi, et al. *Biochimica et Biophysica Acta (BBA)-General Subjects* 1850.5 (2015): 872-877.



Raise, then lower, the temperature
- *Simulated Annealing*

Bernardi, et al. *Biochimica et Biophysica Acta (BBA)-General Subjects* 1850.5 (2015): 872-877.

All atom and Coarse Grained simulation systems

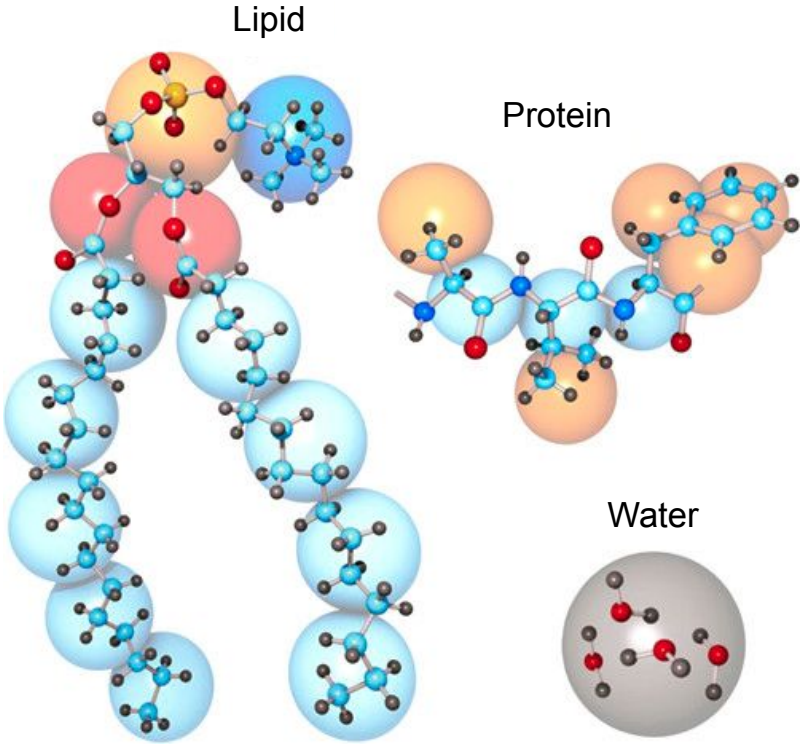


Image from: Kmiecik, S., et al. "Coarse-grained protein models and their applications." *Chemical reviews* 116.14 (2016): 7898-7936.

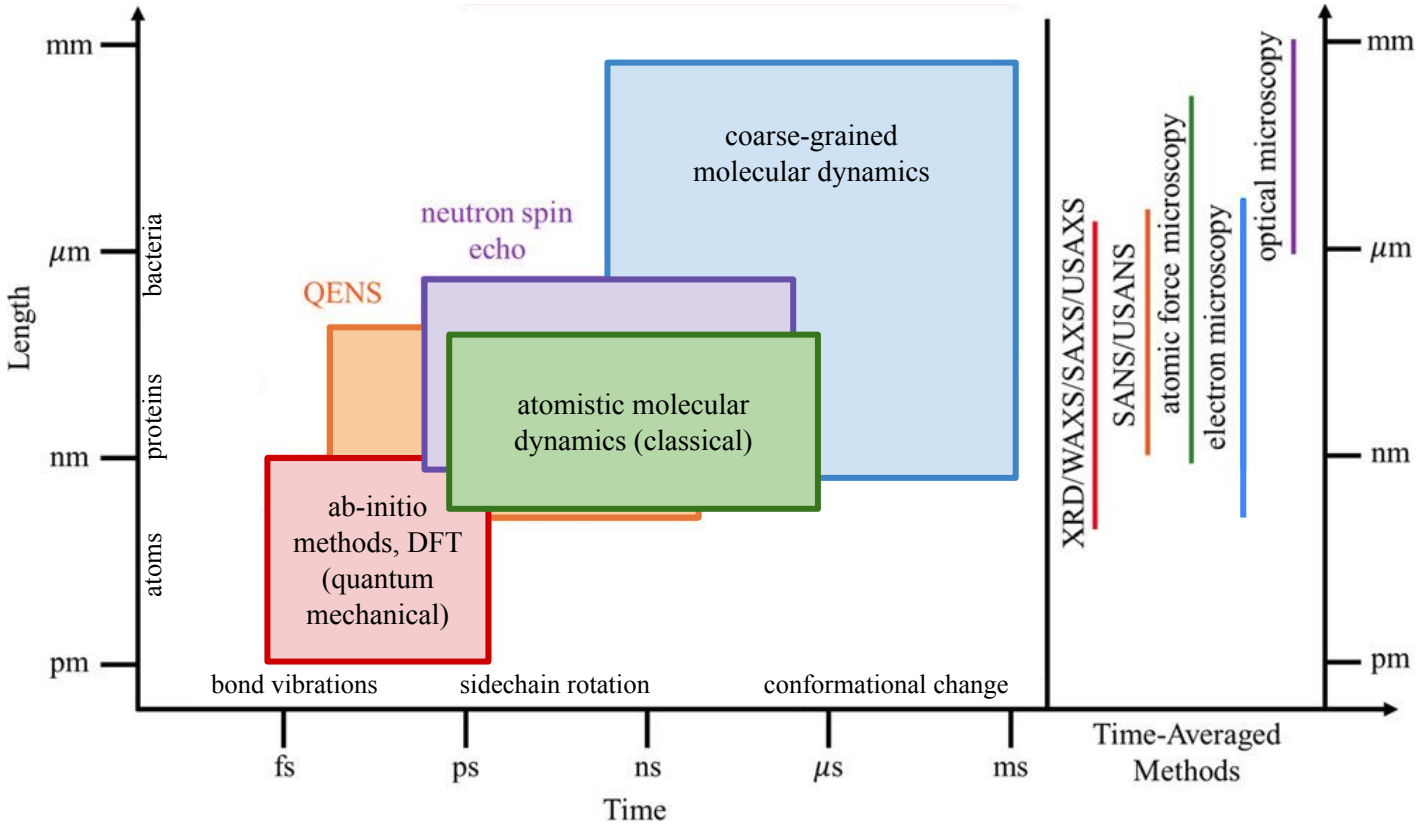
All atom

- Each atom is individually represented
- Captures local motions in detail
- Quickly become computationally expensive with increased system size
    - To enable longer timesteps (i.e. faster simulations), hydrogens are often made heavier or restrained

Coarse grained

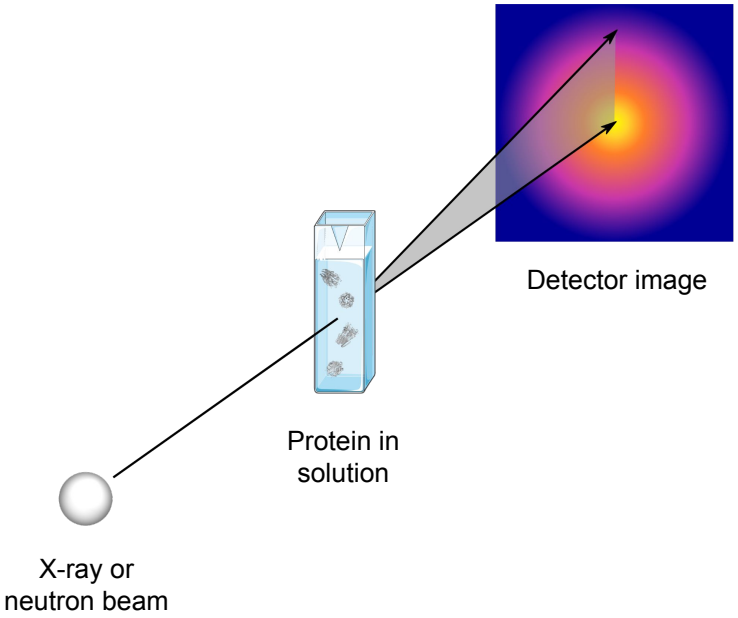
- A few atoms are together represented as a bead
- Enables larger systems or longer timescales to be simulated
- Restraints are often applied to the protein conformation to preserve it
- “Backmapping” is building an all atom model from a coarse grained model
    - Calculating theoretical scattering curves is typically done on all atom models

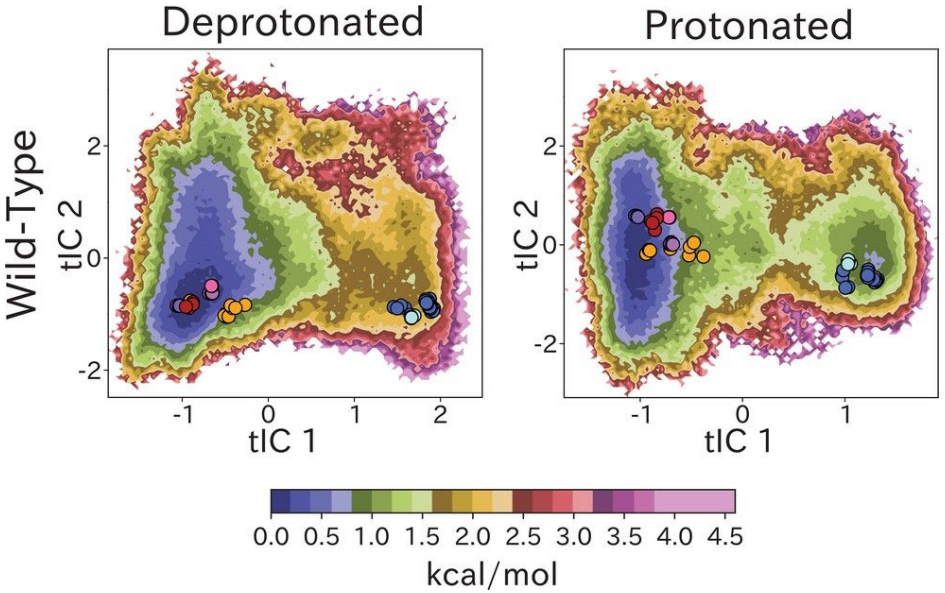
Comparison of timescales and lengthscales of techniques



Adapted from: Wolf, C. M., et al. "Strategies for the Development of Conjugated Polymer Molecular Dynamics Force Fields Validated with Neutron and X-ray Scattering." *ACS Polymers Au* 1.3 (2021): 134-152.

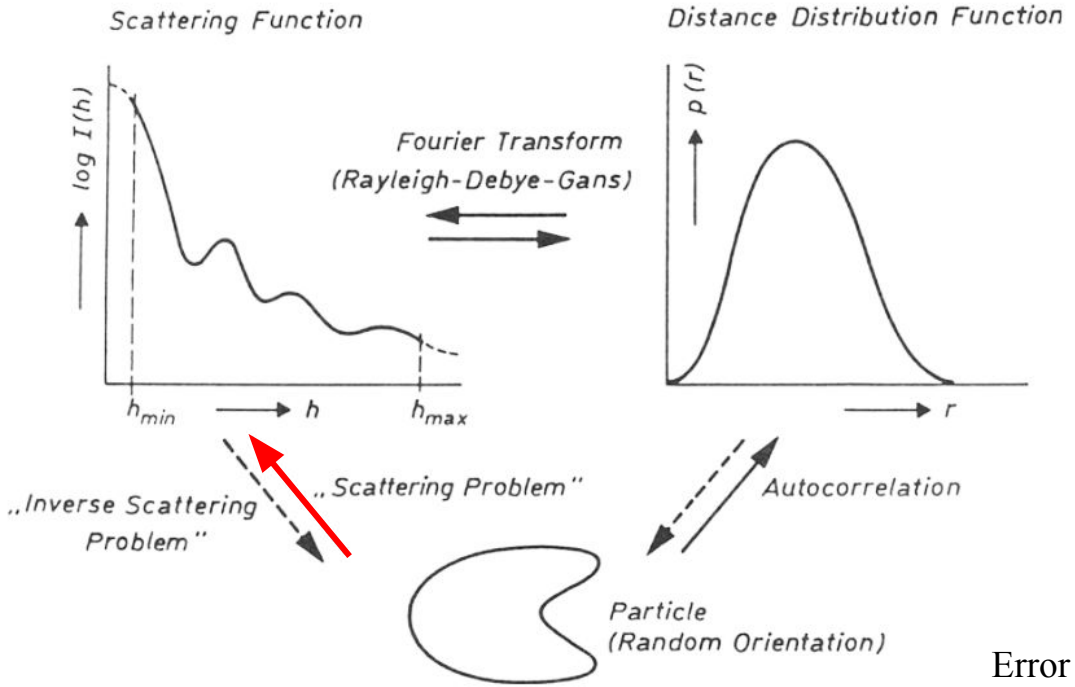
Both SAS and MD simulations can get at the population





Bergh, C., et al. "Markov state models of proton-and pore-dependent activation in a pentameric ligand-gated ion channel." *Elife* 10 (2021): e68369.

Theoretical scattering profiles and comparing to SAS data



Glatter, O. *Modern aspects of small-angle scattering*. Dordrecht: Springer Netherlands, 1995. 107-180.

Software for theoretical scattering profiles:

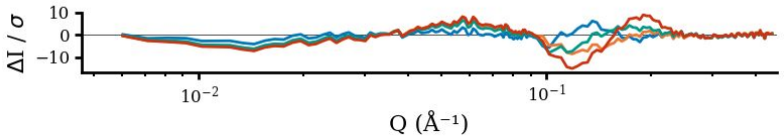
- CRYSOL / CRYSON
- Pepsi-SAXS / Pepsi-SANS
- FoXs

The reduced χ^2 goodness of fit

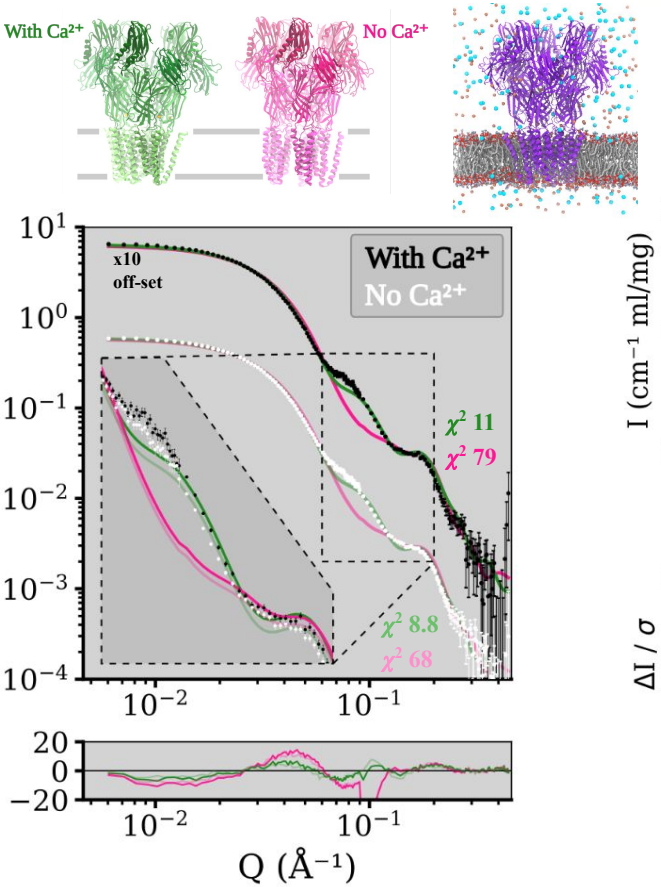
$$\chi^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\frac{I_{model}(q_i) - I_{exp}(q_i)}{\sigma_{exp}(q_i)}\right]^2$$

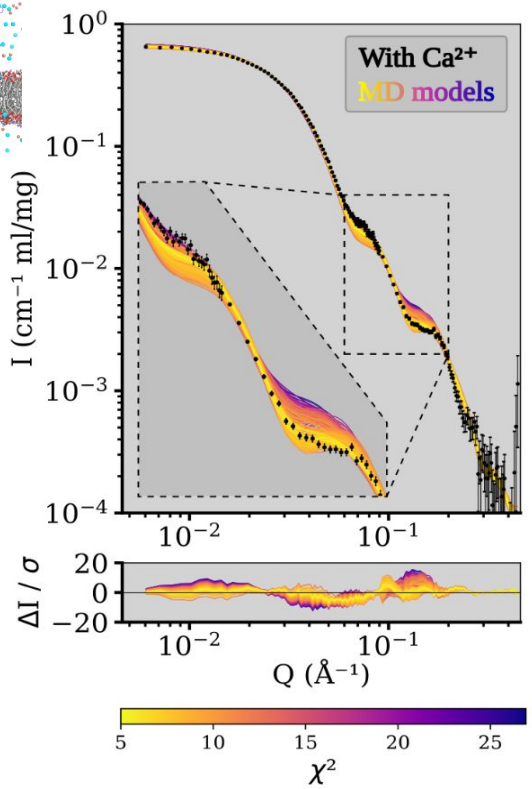
N	number of experimental data points
$I_{model}(q_i)$	calculated intensity of the model at the i-th value of q
$I_{exp}(q_i)$	experimental intensity at the i-th value of q
$\sigma(q_i)$	uncertainty for experimental intensity at i-th q value

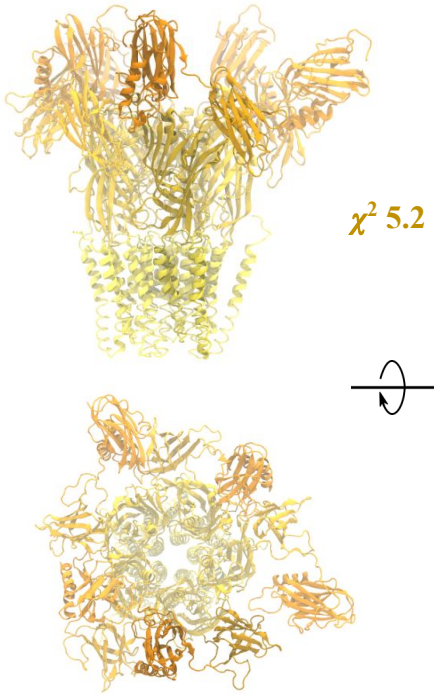
Error weighed residual $\frac{I_{model}(q_i) - I_{exp}(q_i)}{\sigma(q_i)}$



Representative models from MD simulation snapshots



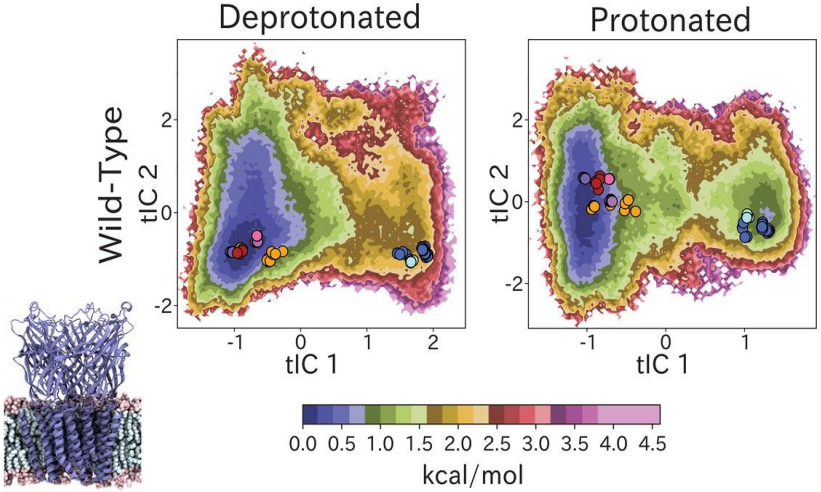




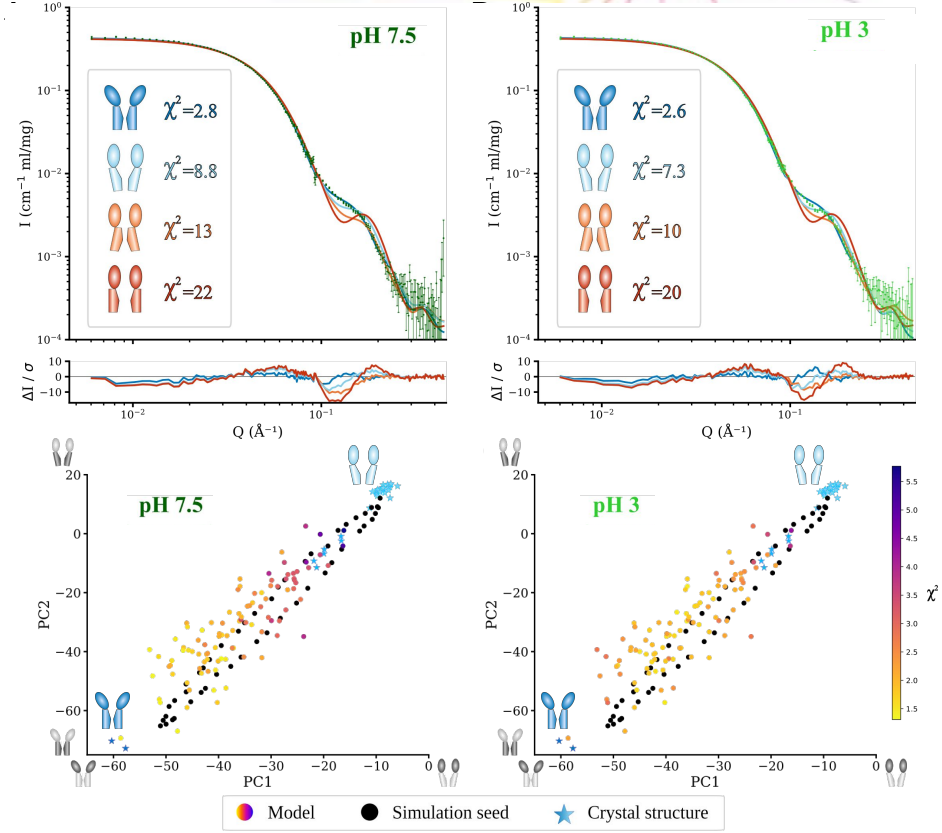
Lycksell, M., Rovšnik, U. et al. "Biophysical characterization of calcium-binding and modulatory-domain dynamics in a pentameric ligand-gated ion channel." *PNAS* 119.50 (2022): e2210669119.

Beyond representative models: Multiple coexisting states

- Finding a model describing the population average is best suited for systems where there is a single dominant population
- For systems with multiple coexisting states, ensemble approaches are more suited



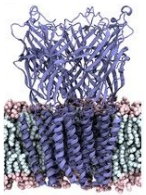
Bergh, C., et al. "Markov state models of proton-and pore-dependent activation in a pentameric ligand-gated ion channel." *Elife* 10 (2021): e68369.

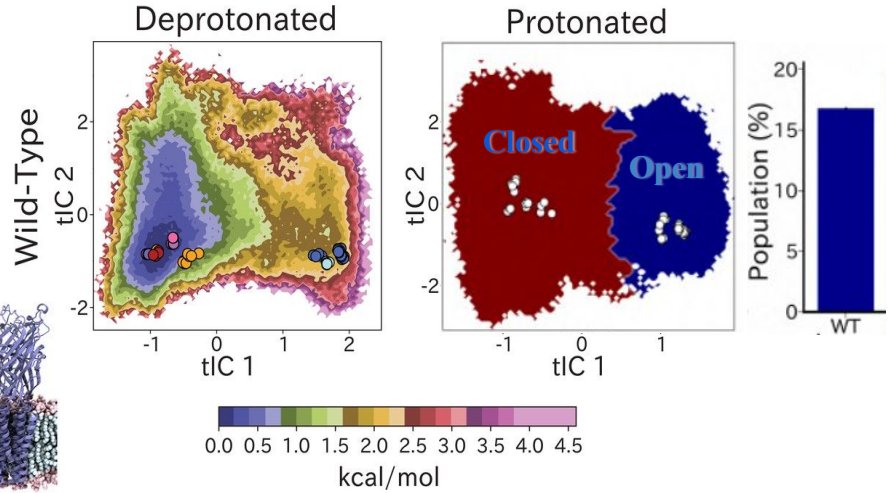


Lycksell, M., et al. "Probing solution structure of the pentameric ligand-gated ion channel GLIC by small-angle neutron scattering." *PNAS* 118.37 (2021): e2108006118.

Beyond representative models: Multiple coexisting states

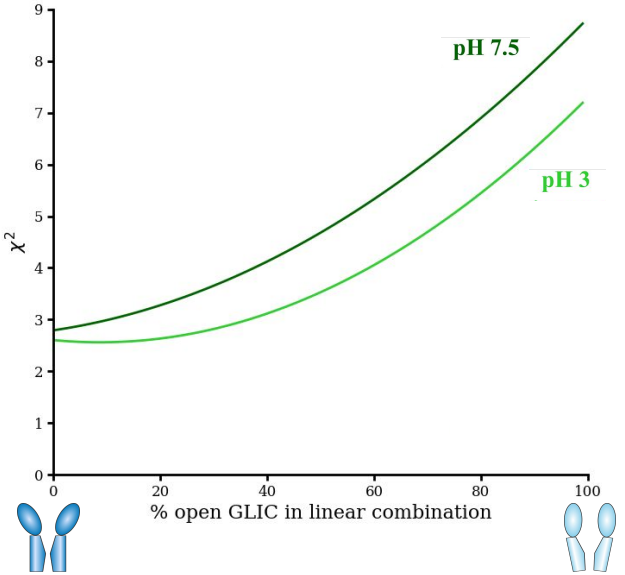
- Finding a model describing the population average is best suited for systems where there is a single dominant population
- For systems with multiple coexisting states, ensemble approaches are more suited





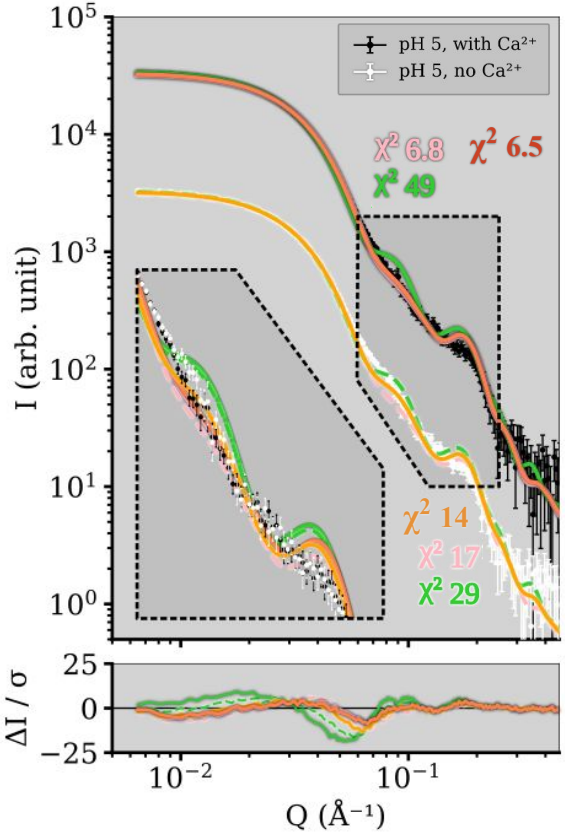
Bergh, C., et al. "Markov state models of proton-and pore-dependent activation in a pentameric ligand-gated ion channel." *Elife* 10 (2021): e68369.

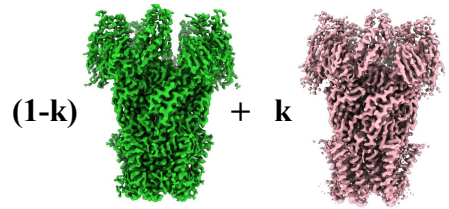
Linear combination: **(1-k)** + **k**

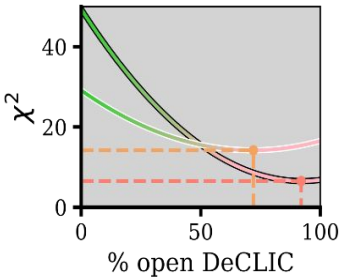


Lycksell, M., et al. "Probing solution structure of the pentameric ligand-gated ion channel GLIC by small-angle neutron scattering." *PNAS* 118.37 (2021): e2108006118.

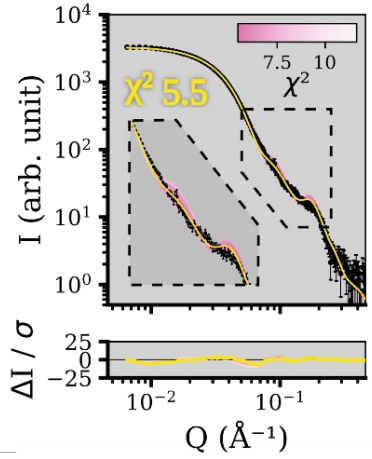
A challenging to fit system

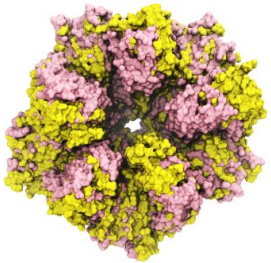






with Ca^{2+}	8%	+	92%	$\Rightarrow X^2 = 6.5$
no Ca^{2+}	28%	+	72%	$\Rightarrow X^2 = 14$





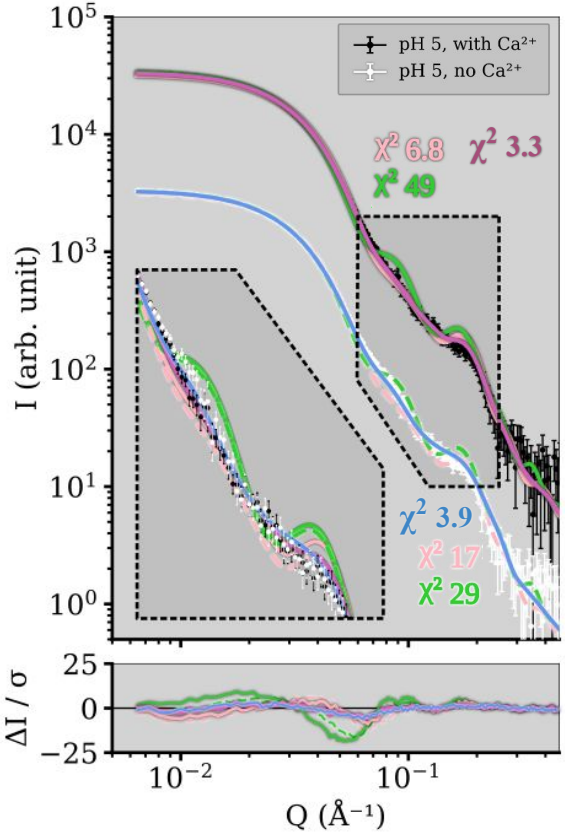
Top view

Starting model

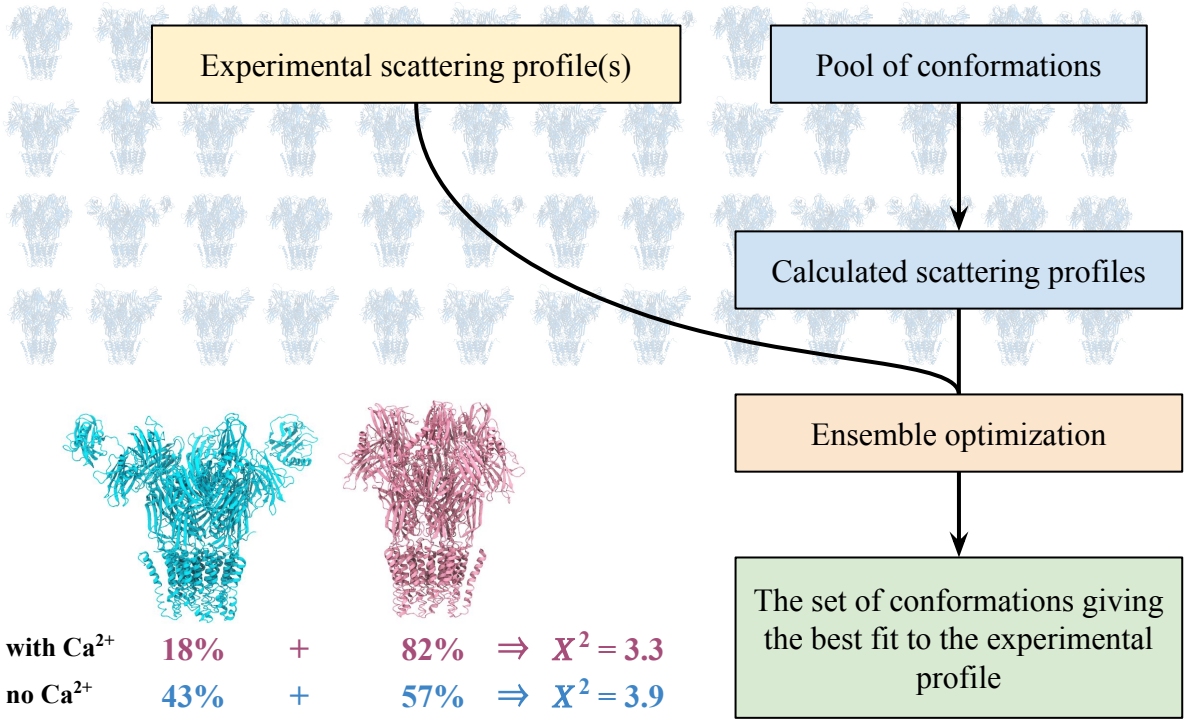
Best fitting snapshot

Unpublished data

Ensemble optimization to choose a set of models

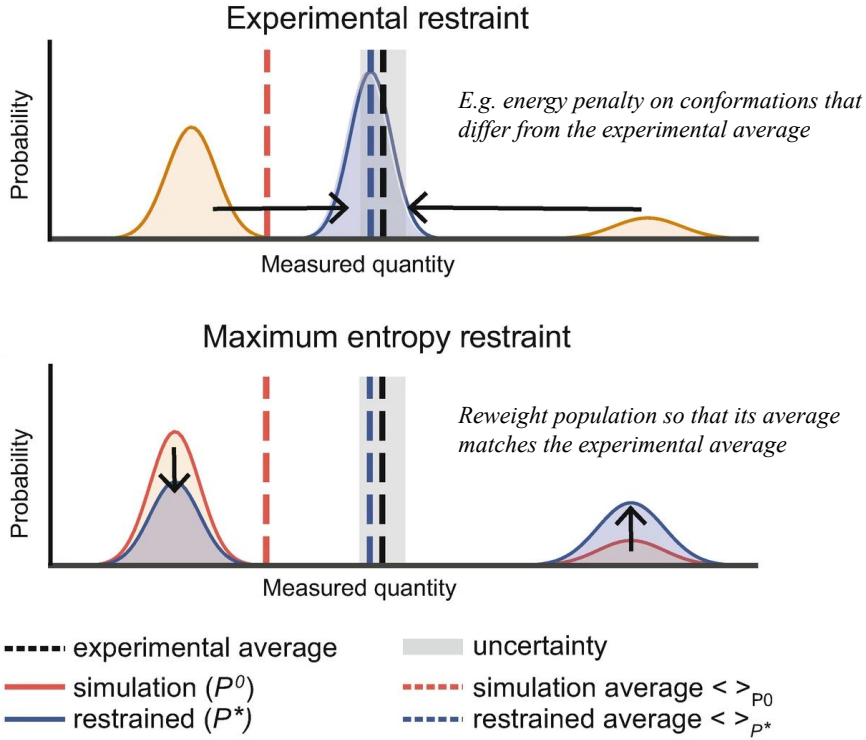


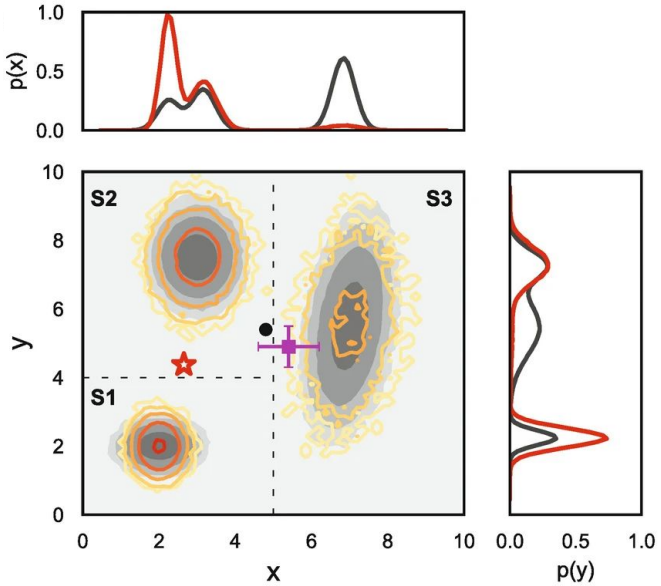
Unpublished data



The EOM software: Bernadó, P., et al. "Structural characterization of flexible proteins using small-angle X-ray scattering." *Journal of the American Chemical Society* 129.17 (2007): 5656-5664.

Biasing simulations vs reweighting an ensemble

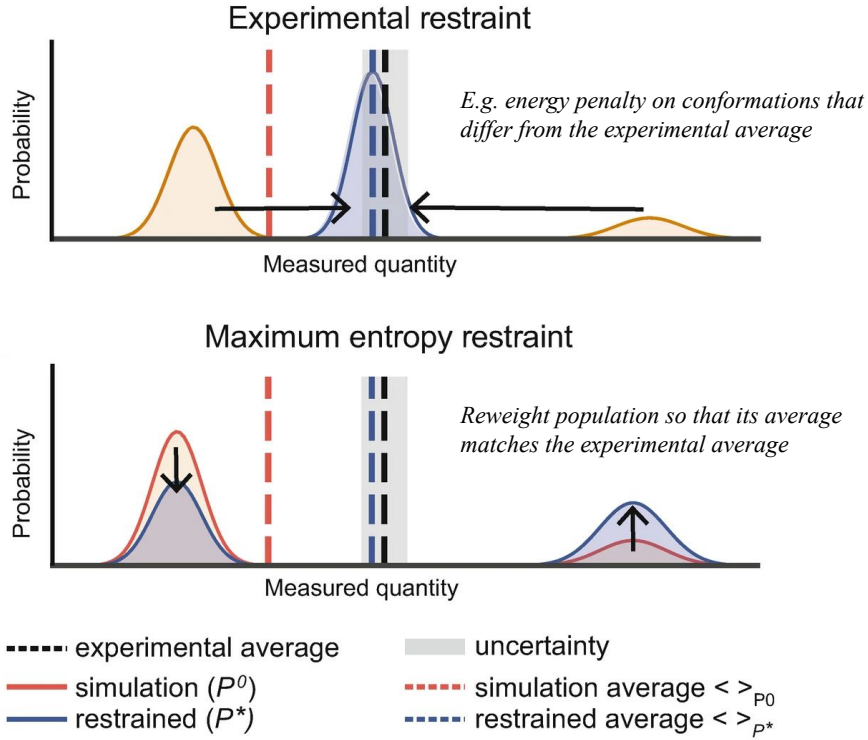


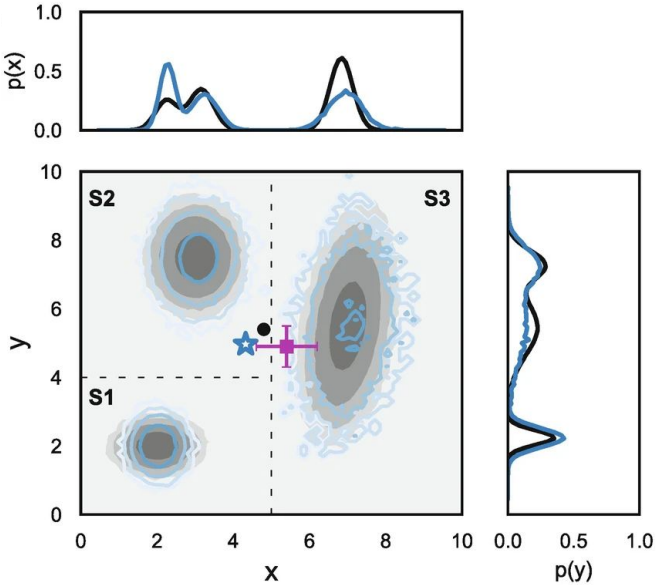


True probability
- True average
- Hypothetical average with uncertainty
- Simulation
- ☆ Simulation average

Bottaro, S., et al. "Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach." *Structural bioinformatics: methods and protocols* (2020): 219-240.

Biasing simulations vs reweighting an ensemble

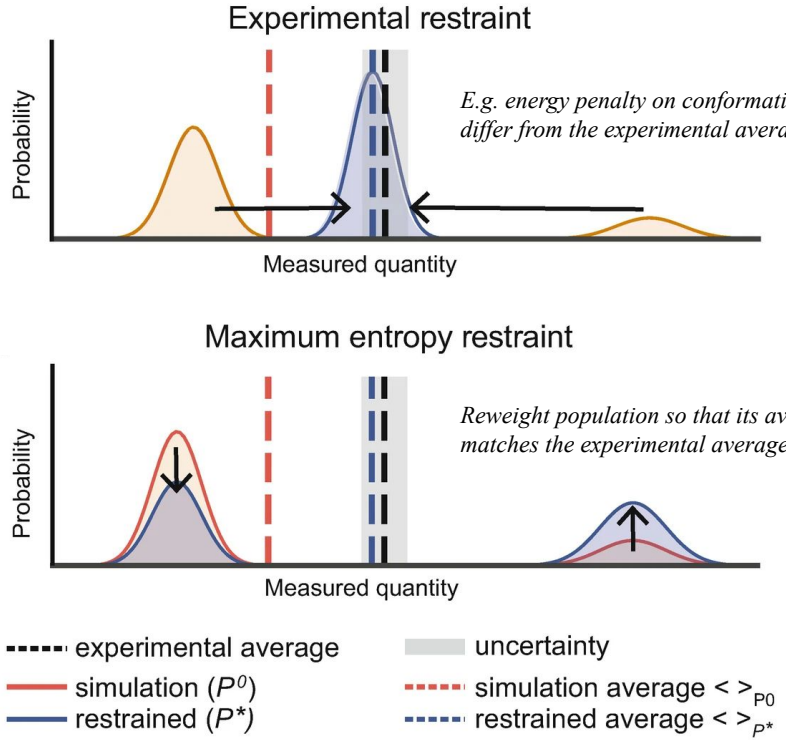


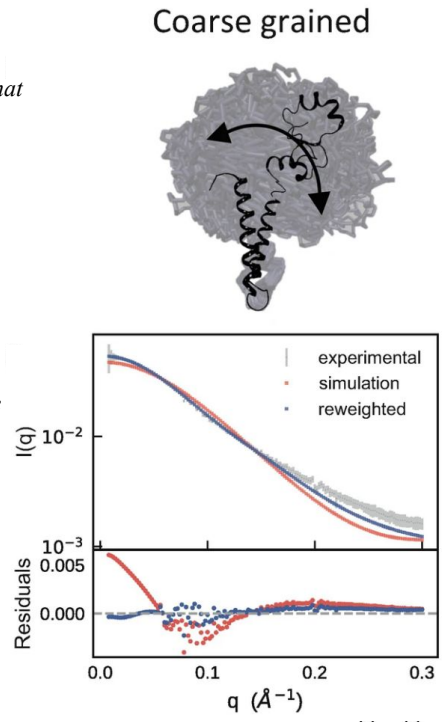


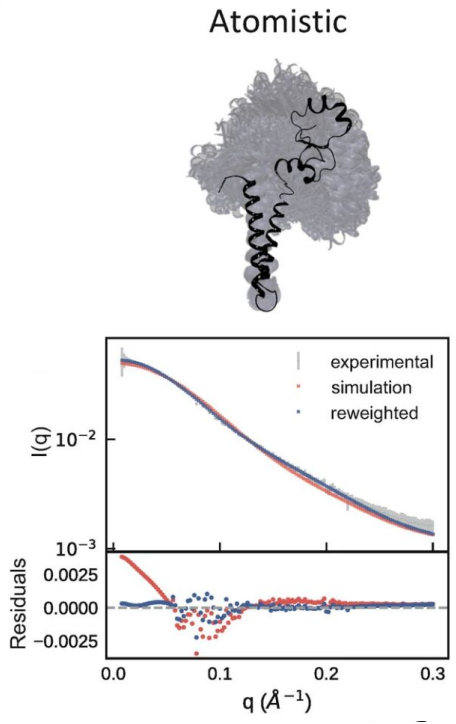
True probability
- True average
- Hypothetical average with uncertainty
- Reweighted simulation
- Simulation average

Bottaro, S., et al. "Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach." *Structural bioinformatics: methods and protocols* (2020): 219-240.

Biasing simulations vs reweighting an ensemble



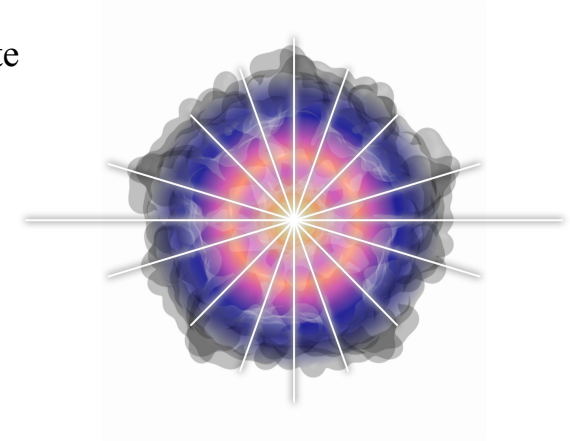




Bottaro, S., et al. "Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach." *Structural bioinformatics: methods and protocols* (2020): 219-240.

Conclusion

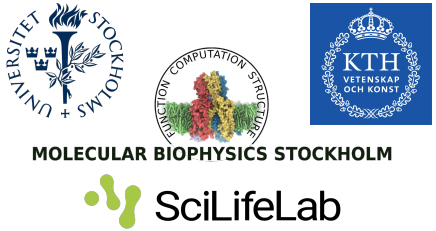
- Molecular dynamics simulation can be a powerful tool in studying a system
  - Requires a sufficiently detailed description of the system to start from
  - Sampling may be an issue, but there are ways to address that should it be needed
- Conformations sampled by MD are physically plausible
  - Ensemble optimization or probability reweighting won't create unreasonable conformations to force a good fit, nor will they create new conformations and are thus limited by the sampling provided to them
- Fitting SAS data with MD simulations can yield improved fits compared to fitting with experimentally determined structures
- MD simulations can help interpret SAS data, and SAS data can corroborate observations from MD simulations
  - E.g. estimates of relative populations
- Which level of detail that is appropriate depends on the system and your goals



Acknowledgements

Anne Martel

Magnus P Andersson



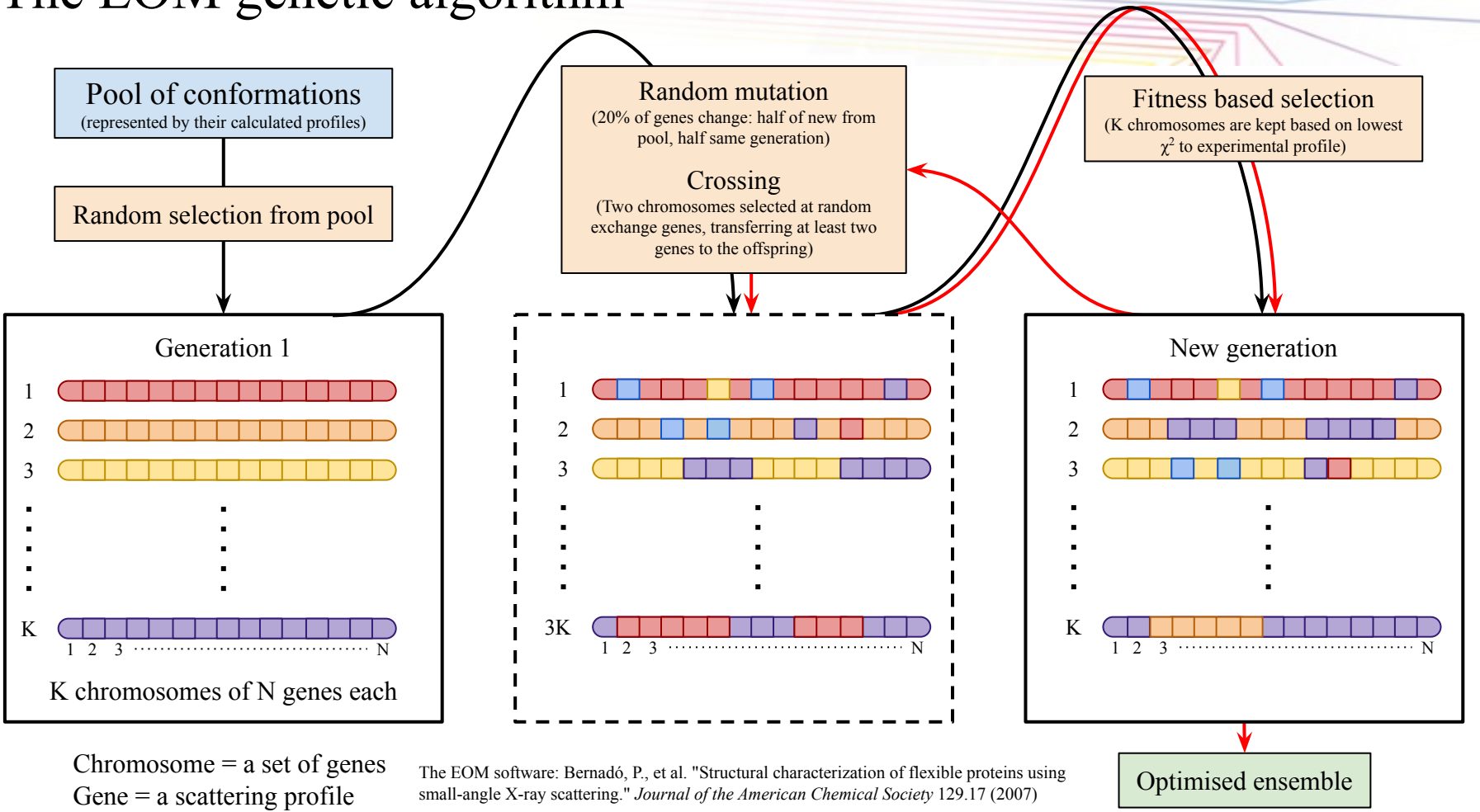
Erik Lindahl
Rebecca Howard
Urška Rovšnik
Cathrine Bergh

The Gromacs Development team

PDC

PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

The EOM genetic algorithm



Chromosome = a set of genes
Gene = a scattering profile

The EOM software: Bernadó, P., et al. "Structural characterization of flexible proteins using small-angle X-ray scattering." *Journal of the American Chemical Society* 129.17 (2007)